Modelling Internal Dynamic Behaviour of BDI Agents[+]

Frances Brazier[a], Barbara Dunin-Keplicz[b], Jan Treur[a], Rineke Verbrugge[a]

[a] Vrije Universiteit Amsterdam
Department of Mathematics and Computer Science, Artificial Intelligence Group
De Boelelaan 1081a, 1081 HV Amsterdam, The Netherlands
Emails: {frances,treur,rineke}@cs.vu.nl
URL: http://www.cs.vu.nl

[b] Warsaw University
Institute of Informatics, ul. Banacha 2, 02-097 Warsaw, Poland
Email: keplicz@mimuw.edu.pl

Abstract
A generic model for the internal dynamic behaviour of BDI agents is proposed. This model, a refinement of a generic agent model, explicitly specifies beliefs and motivational attitudes such as desires, goals, intentions, commitments, and plans, and their relations. A formal meta-language is used to represent beliefs, motivational attitudes and strategies. Dynamic aspects of reasoning about and revision of beliefs and motivational attitudes are modelled in a compositional manner within the modelling framework DESIRE.

1 Introduction

In the last five years multi-agent systems have been a major focus of research in AI. The concept of agents, in particular the role of agents as participants in multi-agent systems, has been subject to discussion. In (Wooldridge and Jennings, 1995) different notions of strong and weak agency are presented. In other contexts big and small agents have been distinguished (Velde and Perram, 1996). In this paper, a model for a rational agent is proposed: a rational agent described using cognitive notions such as beliefs, desires and intentions.

Beliefs, intentions, and commitments play a crucial role in determining how rational agents will act. Shoham defines an agent to be "an entity whose state is viewed as consisting of mental components such as beliefs, capabilities, choices, and commitments. (...) What makes any hardware or software component an agent is precisely the fact that one has chosen to analyze and control it in these mental terms" (Shoham, 1993). This definition provides a basis to study, model and specify mental attitudes; see (Rao and Georgeff, 1991; Cohen and Levesque, 1990; Shoham, 1991; Dunin-Keplicz and Verbrugge, 1996).

[+] In: P.Y. Schobbens and A. Cesta (eds.), Proc. Third International Workshop on Formal Models of Agents, MODELAGE'97, Lecture Notes in AI, Springer Verlag, 1997, in press

The goal of this paper is to define a generic BDI agent model in the compositional multi-agent modelling framework DESIRE. To this purpose, a generic agent model is presented and refined to incorporate beliefs, desires and intentions (in which intentions with respect to goals are distinguished from intentions with respect to plans). The result is a more specific BDI agent in which dependencies between beliefs, desires and intentions are made explicit. The BDI model includes knowledge of different intention/commitment strategies in which these dependencies are used to reason about beliefs, desires, and intentions, but also to explicitly revise specific beliefs, desires and intentions.

The main emphasis in this paper is on static and dynamic relations between mental attitudes. DESIRE (framework for DEsign and Specification of Interacting REasoning components) is a framework for modelling, specifying and implementing multi-agent systems, see (Brazier, Dunin-Keplicz, Jennings, and Treur, 1995, 1996; Dunin-Keplicz and Treur, 1995). Within the framework, complex processes are designed as compositional models consisting of interacting task-based hierarchically structured components. Agents are modelled as composed components. The interaction between components, and between components and the external world, is explicitly specified. Components may be primitive reasoning components using a knowledge base, but may also be subsystems capable of performing tasks using methods as diverse as decision theory, neural networks, and genetic algorithms.

In this paper a small, simplified part of an application, namely meeting scheduling, is used to illustrate the way in which dependencies and strategies are used to model revision.

The paper is structured in the following manner. In Section 2, a generic classification of mental attitudes is presented and a more precise characterization of a few selected motivational attitudes is given. Next, in Section 3, the specification framework DESIRE for multi-agent systems is characterized. In Section 4 a general agent model is described. The framework of modelling motivational attitudes in DESIRE is discussed in Section 5. In Section 6 the use of the explicit knowledge of dependencies and strategies for belief, intention and commitment revision is explained. Finally, Section 7 presents some conclusions and possible directions for further research.

2 Intention and commitment strategies

A number of motivational attitudes, and the static and dynamic relations between motivational attitudes and agents' activities, are modelled in this paper. Individual agents are assumed to have intentions and commitments both with respect to goals and with respect to plans. Joint motivational attitudes and joint actions are not discussed in this paper. The following classification of an agent's attitudes is used:

1. Informational attitudes
   - 1.1 Knowledge
   - 1.2 Beliefs

2. Motivational attitudes
   - 2.1 Desires
   - 2.2 Intentions
     - 2.2.a Intended goals
     - 2.2.b Intended plans

2.3 Commitments
2.3.a Committed goals
2.3.b Committed plans

In this classification the weakest motivational attitude is desire. Desires may be ordered according to preferences and they are the only motivational attitudes subject to inconsistency. A limited number of intended goals are chosen by an agent, on the basis of its (beliefs and) desires. In this paper only achievement goals (and not, for example, maintenance goals) are considered. Moreover, agents are assumed to assure consistency of intentions. With respect to intentions, the conditions elaborated in (Bratman, 1987; Cohen and Levesque, 1990) are adopted.

On the basis of intentions, an agent commits to itself to achieve both goals and to execute plans. In addition an agent may also make commitments to other agents. Such social commitments (Castelfranchi, 1995; Dunin-Keplicz and Verbrugge, 1996) are also explicitly modelled. As proposed in (Castelfranchi, 1995), contrary to some other approaches, social commitments are stronger than intentions, because the aspects of obligation and of interest in the commitment by the other agent are involved.

After committing to a goal and an associated plan, an agent starts plan realization. Knowledge of strategies and dependencies is required to determine in which situations an agent drops an intention or commitment, and how. The kind of behavior that agents manifest depends on immanent behavioral characteristics and environment, including their intention and commitment strategies. As a result individual agents may behave differently in analogical situations. In (Rao and Georgeff 1991) intention strategies were introduced, which inspired the definition of social commitment strategies in (Dunin-Keplicz and Verbrugge, 1996). These commitment strategies include the additional aspects of communication and coordination.

In this paper, three commitment strategies are distinguished. The strongest commitment strategy is followed by the *blindly committed* agent, that maintains its commitments until it believes they have been achieved, irrespective of changes in its own goals and desires, and irrespective of other beliefs with respect to the feasibility of the commitment. A *single-minded* agent may drop commitments when it believes they can no longer be attained, irrespective of changes in its goals and desires. However, as soon as a single-minded agent abandons a commitment, communication and coordination are necessary with agents to whom the single-minded agent is committed. An *open-minded* agent may drop commitments when it believes they can no longer be attained or when the relevant goals are no longer desired. Communication and coordination with agents to whom the single-minded agent is committed, are also performed when commitments are abandoned.

For simplicity, in this paper each agent is assumed to follow a single commitment strategy during the whole process of plan realization. Moreover, it should be stressed that commitment strategies are used for both committed goals and committed plans.

3 A modelling framework for Multi-Agent Systems

The compositional BDI model introduced in this paper is based on an analysis of the tasks performed by a BDI agent. Such a task analysis results, among others, in a (hierarchical) task composition, which is the basis for a compositional model: components in a compositional model are directly related to tasks in a task

composition. Interaction between tasks is modelled and specified at each level within a task composition, making it possible to explicitly model tasks which entail interaction between agents. The hierarchical structures of tasks, interaction and knowledge are fully preserved within compositional models. Task coordination is of importance both within and between agents. Below the formal compositional framework for modelling multi-agent tasks DESIRE is briefly introduced, in which the following aspects are modelled and specified (for more details, see (Brazier, Dunin-Keplicz, Jennings, Treur, 1997)):

(1) a task composition,
(2) information exchange,
(3) sequencing of tasks,
(4) task delegation,
(5) knowledge structures.

3.1 Task composition

To model and specify composition of tasks, knowledge of the following types is required:

- a *task hierarchy*,
- information a task requires as *input*,
- information a task produces as a *result* of task performance
- *meta-object* relations between tasks

Within a task hierarchy *composed* and *primitive* tasks are distinguished: in contrast to primitive tasks, composed tasks consist of a number of other tasks, which, in turn, may be either composed or primitive. Tasks are directly related to components: composed tasks are specified as composed components and primitive tasks as primitive components.

Information required/produced by a task is defined by *input* and *output signatures* of a component. The signatures used to name the information are defined in a predicate logic with a hierarchically ordered sort structure (order-sorted predicate logic). Units of information are represented by the ground *atoms* defined in the signature.

The role information plays within reasoning is indicated by the level of an atom within a signature: different (meta)levels may be distinguished. In a two-level situation the lowest level is termed *object-level information*, and the second level *meta-level information*. Meta-level information contains information about object-level information and reasoning processes; for example, for which atoms the values are still unknown (*epistemic information*). Similarly, tasks which include reasoning about other tasks are modelled as meta-level tasks with respect to object-level tasks. Often more than two levels of information and reasoning occur, resulting in meta-meta-... information and reasoning.

3.2 Information exchange between tasks

Information links between components are used to specify information exchange between tasks. Two types of information links are distinguished: *private* information

links and *mediating* information links. For a given parent component, a private information link relates output of one of its components to input of another, by specifying which truth value of a specific output atom is linked with which truth value of a specific input atom. Atoms can be renamed: each component can be specified in its own language, independent of other components. In a similar manner mediating links transfer information from the input interface of the parent component to the input interface of one of its components, or from the output interface of one of its components to the output interface of the parent component iteself. Mediating links specify the relation between the information at two adjacent levels in the component hierarchy. The conditions for activation of information links are explicitly specified as task control knowledge.

3.3 Sequencing of tasks

Task sequencing is explicitly modelled within components as *task control knowledge*. Task control knowledge includes not only knowledge of which tasks should be activated, when and how, but also knowledge of the goals associated with task activation and the extent to which goals should be derived. These aspects are specified as component and link activation together with task control foci and extent to define the component's goals. Components are, in principle, black boxes to the task control of an encompassing component: task control is based purely on information about the success and/or failure of component reasoning. Reasoning of a component is considered to have been successful with respect to an evaluation criterion if it has reached the goals specified by this evaluation criterion to the extent specified (e.g., any or every).

3.4 Delegation of tasks

During knowledge acquisition a task as a whole is modelled. In the course of the modelling process decisions are made as to which tasks are (to be) performed by which agent. This process, which may also be performed at run-time, results in the delegation of tasks to the parties involved in task execution. In addition to these specific tasks, often generic agent tasks, such as interaction with the world (observation) and other agents (communication and cooperation) are assigned.

3.5 Knowledge structures

During knowledge acquisition an appropriate structure for domain knowledge must be devised. The meaning of the concepts used to describe a domain and the relations between concepts and groups of concepts, are determined. Concepts are required to identify objects distinguished in a domain (domain-oriented ontology) , but also to express the methods and strategies employed to perform a task (task-oriented ontology). Concepts and relations between concepts are defined in hierarchies and rules based on order-sorted predicate logic. In a specification document references to appropriate knowledge structures (specified elsewhere) suffice; compositional knowledge structures are composed by reference to other knowledge structures.

4 Global structure of a generic agent

To model an agent capable of reasoning about its own tasks, processes and plans, its knowledge of other agents, its communication with other agents, its knowledge of the world and its interaction with the world, a generic agent architecture has been devised in which such types of reasoning are transparently allocated to specific components of an agent (see (Brazier, Jonker and Treur, 1997)).

This generic architecture can be applied to different types of agents. In this paper this architecture is refined to model a rational agent with motivational attitudes: other architectures are more applicable for other types of agents. The generic architecture is described in this section, while the refined BDI architecture is the subject of Section 5.

Four of the five types of knowledge distinguished above in Section 3 are used to describe this generic architecture: task composition, information exchange, sequencing of tasks and knowledge structures. Within an individual agent, task delegation is trivial.

4.1 Task composition

As stated above an agent needs to be capable of reasoning about its own processes, its own tasks, other agents and the world. In other words, an agent needs to be capable of six tasks:

(1) controlling its own processes,
(2) performing its own specific tasks,
(3) managing its interaction with the world (observation, execution of actions),
(4) managing its communication with other agents,
(5) maintaining information on the world, and
(6) maintaining information on other agents.

4.2 Information exchange

Information links are defined for the purpose of information exchange between components. The component agent_interaction_management receives information from, and sends information to, other agents. The component world_interaction_management on the other hand exchanges information with the external world. Both components also exchange information with the component own_process_control. Which information is required by an agent specific task depends on the task itself and therefore cannot be predefined. To fully specify the exchange of information, a more specific analysis of the types of information exchange is required. In Figure 1, a number of information links defined for information exchange at the top level of the agent, are shown together with the names of the components they connect.

Link name	*From component*	*To component*
import_world_info	agent (input interface)	world_interaction_management
export_world_info	world_interaction_management	agent (output interface)
transfer_comm_world_info	agent_interaction_management	maintenance_of_world_information
provide_world_state_info	world_interaction_management	own_process_control
import_agent_info	agent (input interface)	agent_interaction_management
export_planned_comm	agent_interaction_management	agent (output interface)
provide_agent_info	agent_interaction_management	own_process_control
transfer_committed_acts&obs	own_process_control	world_interaction_management
transfer_agent_commitments	own_process_control	agent_interaction_management
transfer_planned_comm	own_process_control	agent_interaction_management

Figure 1 Links for information exchange at the top level of an agent

In Figure 2 a graphical representation of the generic architecture for an agent is shown; in this figure a number of the information links and the components they connect, are depicted.

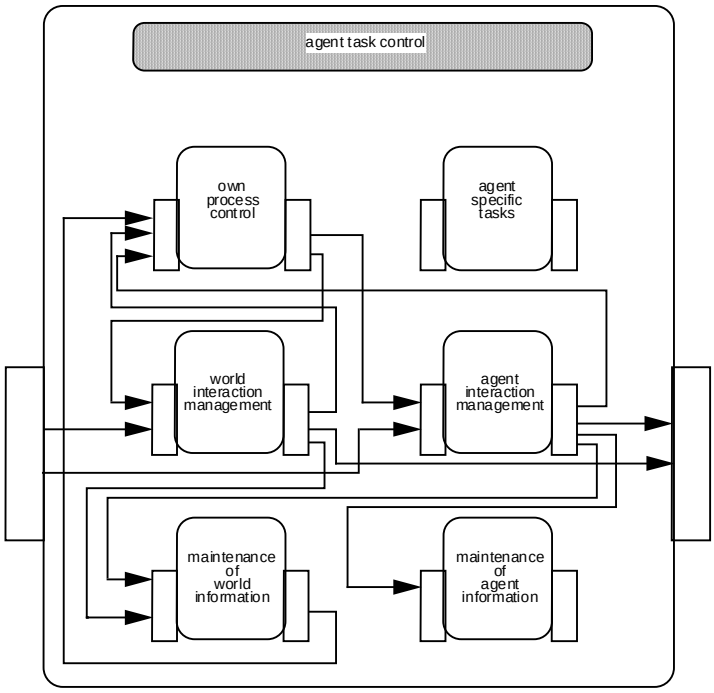

Figure 2 Top level composition and information links of a generic agent

4.3 Task sequencing

Minimal task control has been modelled and specified for the top level of the generic agent. Task control knowledge specifies that all generic components and links are initially awakened. The awake status specifies that as soon as new information arrives, it is processed. This allows for parallel processing of information by different components. The links which connect an agent to other agents are activated by the agents from which they originate. Global task control includes specifications such as the following rule:

if start
then next_component_state(own_process_control, awake)
and next_component_state(world_interaction_management, awake)
and next_component_state(agent_interaction_management, awake)
and next_link_state(import_agent_info, awake)
and next_link_state(export_agent_info, awake)
and next_link_state(import_world_info, awake)
and next_link_state(export_world_info, awake)
and next_link_state(transfer_comm_world_info, awake)
.....

4.4 Knowledge structures

Generic knowledge structures are used within the specification of a generic agent, a number of which have been shown above. In the following section more detailed examples of specifications of knowledge structures will be shown for a rational agent with motivational attitudes.

4.5 Building a real agent

Each of the six components of the generic agent model presented above can be refined in many ways, resulting in models of agents with different characteristics. (Brazier, Jonker and Treur, 1996) describe a model of a generic cooperative agent, based on the generic agent model and Jenning's model of cooperation, see (Jennings, 1995). In (Brazier and Treur, 1996) another refinement of the generic agent model is proposed for reflective agents capable of reasoning about their own reasoning processes and other agents' reasoning processes. In the following section a refinement of the component own_process_control is presented in which motivational attitudes (including beliefs, desires and intentions) play an important role.

5 A model for rational agents with motivational attitudes

The generic model and specifications of an agent described above, can be refined to a generic model of a rational BDI agent capable of explicit reasoning about its beliefs, desires, intentions and commitments. First, some of the assumptions behind the model are discussed (Section 5.1). Next the specification of the model is presented for the highest level of abstraction (in Section 5.2 and 5.3), and for the more specific levels of abstraction (Section 5.4).

5.1 Rational agents with motivational attitudes

Before presenting the model, some of the assumptions upon which this model is based, are described. Agents are assumed to be rational: they must be able to generate goals and act rationally to achieve them, namely planning, replanning, and plan execution. Moreover, to fully adhere to the strong notion of agency, an agent's activities are described using mentalistic notions usually applied to humans. This does not imply that computer systems are believed to actually "have" beliefs and intentions, but that these notions are believed to be useful in modelling and specifying the behaviour required to build effective multi-agent systems (see, for example, (Dennett, 1987) for a description of the "intentional stance").

A first assumption is that motivational attitudes, such as beliefs, desires, intentions and commitments are defined as *reflective statements* about the agent itself and about the agent in relation to other agents and the external world. These reflective statements are modelled in DESIRE in a meta-language, which is order sorted predicate logic. Functional or logical relations between motivational attitudes and between motivational attitudes and informational attitudes are expressed as meta-knowledge, which may be used to perform meta-reasoning resulting in further conclusions about motivational attitudes. For example, in a simple instantiation of the model, beliefs can be inferred from meta-knowledge that any observed fact is a believed fact and that any fact communicated by a trustworthy agent is a believed fact.

A second assumption is that information is classified according to its *source*: internal information, observation, communication, deduction, assumption making. Information is explicitly labeled with these sources. Both informational attitudes (such as beliefs) and motivational attitudes (such as desires) depend on these sources of information. Explicit representations of the dependencies between attitudes and their sources are used when update or revision is required.

A third assumption is that the *dynamics* of the processes involved are explicitly modelled. For example, a component may be made awake from the start, which means that it always processes incoming information immediately. If more components are awake, their processes will run in parallel. But, if tasks depend on each other, sequential activation may be preferred. Both parallel and sequential activation may be specified explicitly. If required, update or revision takes place and is propagated through different components by active information links.

A fourth assumption is that the model presented below is *generic*, in the sense that the explicit meta-knowledge required to reason about motivational and informational attitudes has been left unspecified. To tune the model to a given application this knowledge has to be added. In this paper, examples of the types of knowledge are given for the purpose of illustration.

A fifth assumption is that intentions and commitments are defined with respect to both *goals and plans*. An agent accepts commitments towards itself as well as towards others (social commitments). In this paper, an agent determines which goals it intends to fulfill, and commits to a selected subset of these goals. Similarly, an agent determines which plans it intends to perform, and commits to a selected subset of these plans.

Most reasoning about beliefs, desires, and intentions can be modelled as an essential part of the reasoning an agent needs to perform to control its own processes. A refinement of the generic component own_process_control described in Section 4 is presented below.

5.2 A refined model of own process control

Finally, to design a BDI agent, the component own_process_control is refined. The component own_process_control is composed of three components, which reason about:

(1) the agent's beliefs
(2) its desires
(3) its intentions and commitments with respect to both goals and plans.

The extended task hierarchy for a BDI agent is shown in Figure 3. The component belief_determination performs reasoning about relevant beliefs in a given situation. In the component desire_determination an agent determines which desires it has, related to its beliefs. Intended and committed goals and plans are derived by the component intention_and_commitment_determination. This component first determines the goals and/or plans it intends to pursue before committing to the specific selected goals and/or plans. All three components are further refined in Section 5.4.

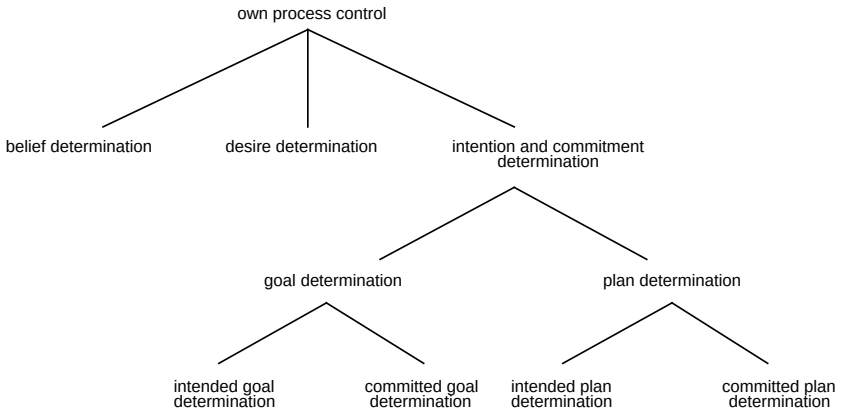

Figure 3 Task hierarchy of own process control within a BDI agent

In the model, beliefs and desires influence each other reciprocally. Furthermore, beliefs and desires both influence intentions and commitments. This is explicitly modelled by information links between the components and meta-knowledge within each of the components.

In Figures 4.1 and 4.2, the composition of own_process_control is shown, together with the exchange of information. This is specified in DESIRE graphically as in Figure 4.1.

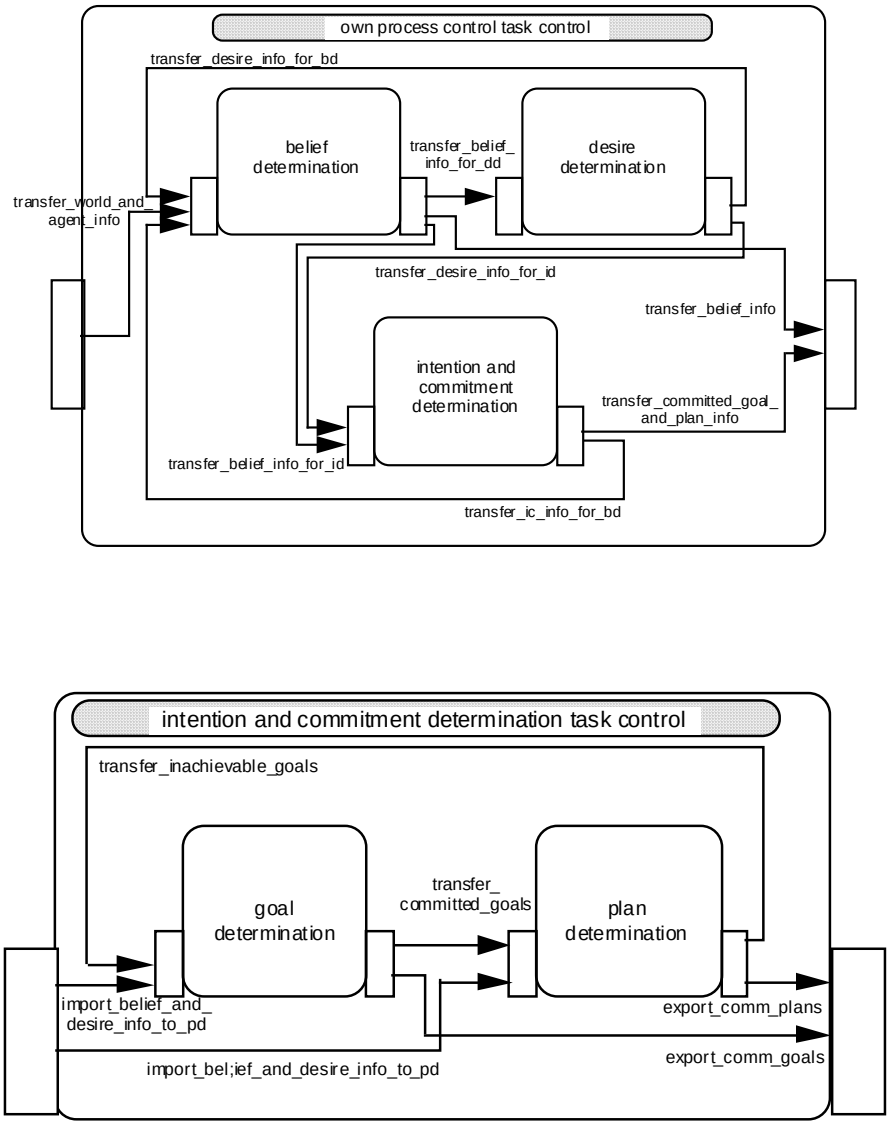

Figure 4.1 Refinement of own process control within the BDI agent

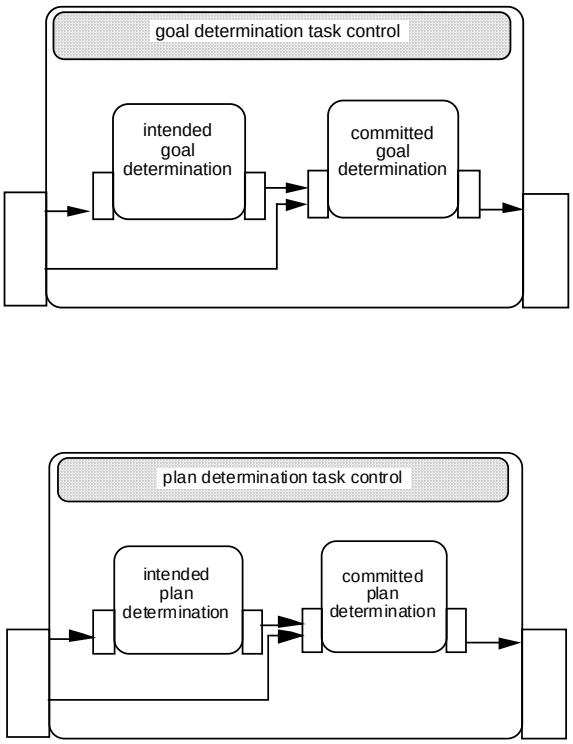

Figure 4.2 Further refinement of goal determination and plan determination

Task control knowledge of the component own_process_control determines that:

(1) initially all links within the component own_process_control are awakened, and the component belief_determination is activated,

(2) once the component belief_determination has succeeded in reaching all possible conclusions (specified in the evaluation criterion goals) desire_determination is activated and belief_determination is made continually active (awake),

(3) once the component desire_determination has succeeded in reaching all possible conclusions (specified in the evaluation criterion desires), the component intention_and commitment_determination is activated and desire_determination is made continually active (awake). In addition, the desires in which the agent may want to believe (wishful thinking) are transferred to the component belief_determination.

Task control of the component intention_and_commitment_determination, in turn, is described in Section 5.4.3.

5.3 The global reasoning strategy

The global reasoning strategy specified by task control knowledge in the model is that some chosen desires (depending on knowledge in the component intended_goal_determination, existing beliefs and specific agent characteristics) become intentions, and some selected intentions (depending on knowledge in the component committed_goal_determination and specific agent characteristics) are translated into committed_goals to the agent itself and to other agents. The agent then reasons about ways to achieve the committed_goals on the basis of knowledge about planning in the component committed_plan_determination, resulting in the construction of a committed_plan. This plan is transferred to one or more of the other high-level components of the agent (depending on the plan in question), namely world_management, agent_management, and agent_specific_tasks, to be executed.

5.4 Further refinement of components

In the previous two sections the model for reasoning about motivational attitudes was described in terms of the three tasks within the component own_process_control and their mutual interaction. In this section each of the tasks themselves is described in more detail.

5.4.1 Belief determination

The task of belief determination requires explicit meta-reasoning to generate beliefs. The specific knowledge used for this purpose obviously depends on the domain of application. The adopted model specifies meta-knowledge about beliefs based on six different sources:

(1) *internal beliefs of an agent*
Internal beliefs are beliefs which an agent inherently has, with no further indication of their source. They can be expressed as meta-facts of the form internal_belief(X:Statement), meaning that X:Statement is an internal belief. These meta-facts can be specified as initial facts or be inferred from other internal meta-information. By meta-knowledge of the form

if internal_belief(X:Statement) **then** belief(X:Statement)

beliefs can be derived from the internal beliefs.

(2) *beliefs based on observations*
Beliefs based on observations are acquired on the basis of observations of the world, either at a particular moment or over time. Simple generic meta-knowledge can be used to derive such beliefs:

if observed_world_fact(X:Statement) **then** belief(X:Statement).

(3) *beliefs based on communication with other agents*
Communication with other agents may, if agents are considered trustworthy, result in beliefs about the world or about other agents. Generic meta-knowledge that can be used to derive such beliefs is:

if communicated_fact_by(X:Statement, A:Agent) **and** trustworthy(A:Agent)
then belief(X:Statement)

(4) *beliefs deduced from other beliefs*
Deduction from other beliefs can be performed by means of an agent's own (domain-dependent) knowledge of the world, of other agents and of itself.

(5) *beliefs based on assumptions*
Beliefs based on assumptions may be derived from other beliefs (and/or from epistemic information on the lack of information) on the basis of default knowledge, knowledge about likelihood, et cetera. For example, a default rule (a : b) / c can be specified as meta-knowledge (e.g. according to the approach described by (Tan and Treur, 1992)).

(6) *beliefs based on desires*
In the case of wishful thinking beliefs may be implied by generated desires. For example, as an extreme case, a strongly wishful-thinking agent may have the following knowledge in belief_determination:

if not belief(not(X:Statement)) **and** desired(X:Statement) **then** belief(X:Statement)

A more sophisticated model to generate beliefs can also keep track of the source of a belief. This can be specified in the meta-language by adding labels to beliefs reflecting their source, for example by belief(X:Statement, L:Label). Here the label L:Label can denote a single source, such as observed, or communicated_by(A:Agent), but if beliefs have been combined to generate other beliefs, also combined labels can be generated as more complex term structures, expressing that a belief depends on a number of sources.

Another aspect of importance is the omniscience problem (Fagin et al., 1995), which requires the control of the belief generation process. In practical reasoning processes, only those beliefs are generated that are of specific interest. Specific solutions to the omniscience problem may be modelled explicitly within this component.

5.4.2 Desire determination

Desires can refer to a (desired) state of affairs in the world (and the other agents), but also to (desired) actions to be performed. Often, desires are influenced by beliefs. Because beliefs can be based on their source, as discussed in Section 5.4.1, desires can inherit these sources. In addition, desires can have their own internal source, for example desires can be inherent to an agent. Knowledge on how desires are generated is left unspecified in the generic model.

5.4.3 Intention and commitment determination

Intended and committed goals and plans are determined by the component intention_and_commitment_determination; this component is composed of the component goal_determination and plan_determination. Each of these two components first determines the intended goals and/or plans it wishes to pursue before committing to a specific goal and/or plan.

In the component goal_determination commitments to goals are generated in two stages. In the component intended_goal_determination, based on beliefs and desires, but also on preferences between goals, specific goals become intended goals. Different agents have different strategies to choose which desires will become intentions. For example:

- some (eager) agents may choose a desire as an intention as soon as it is consistent with their previously established intended goals;
- others (socially complying agents) may select an intention when it is one of their desires which is an intention of other agents with which they automatically comply;
- and still others (apathetic agents) may select no intentions at all.

These differences in agent characteristics can be expressed in the (meta-)knowledge specified for intended_goal_determination. For each intended goal a condition (in the form of **not** inadequate_intended_goal(X:Statement)) is specified that expresses the adequacy of the goal, i.e., that the goal is not subject to revision. As soon as it has been established that the intention has to be dropped, the intended goal becomes inadequate, so this condition no longer holds, which in turn leads to the retraction of the intended goal on the basis of the revision facilities built-in in the semantics and execution environment of DESIRE.

In the component committed_goal_determination a number of intended goals are selected to become goals to which the agent commits; again, different agents have different strategies to select committed goals, and these different strategies can be expressed in the (meta-)knowledge specified for the component committed_goal_determination. The committed goals are transferred to the component plan_determination. In a manner similar to intended goal determination, the knowledge specified for the component committed_goals includes a condition inadequate_committed_goal(X:Statement) that plays a role in revision.

In the component plan_determination commitments to goals are analysed and commitments to plans are generated in two stages. In the component intended_plan_determination plans are generated dynamically, combining primitive actions and predefined plans known to the agent (stored in an implementation, for example, in a library). On the basis of knowledge of the quality of plans, committed goals, beliefs and desires, a number of plans become intended plans. The component committed_plan_determination determines which of these plans should actually be executed. In other words, to which plans an agent commits. If no plan can be devised to reach one or more goals to which an agent has committed, this is made known to the component goal_determination. If a plan has been devised, execution of a plan includes determining, at each point in time, which actions are to be executed. During plan execution, monitoring information can be acquired by the agent through observation and/or communication. Plans can be adapted on the basis of observations and communication, but also on the basis of new information on goals to which an agent has committed. If, for example, the goals for which a certain plan has been

devised, are no longer relevant, and thus withdrawn from an agent's list of committed goals, it may no longer make sense to execute this plan.

6 Modelling commitment strategies

Specifications in DESIRE define in a declarative manner the behaviour of a multi-agent system with respect to their integrated reasoning processes and acting processes (observing, communicating, executing actions in the world). Characteristic to this approach to modelling multi-agent systems is that strategies, revision, and the integration of communication, observation and action in the reasoning process, are explicitly modelled and specified.

6.1 Specification of commitment strategies

After plan construction, the phase of plan realization starts. During this phase, all components of own_process_control are continually awake, so that any revision of an agent's informational and motivational attitudes is propagated immediately by transfer of the new information through links to other components. The fact that both information links and components are always awake ensures that this happens without further explicit specification of activation. Thus, new information is not necessarily expected at specific points in the process.

In our model, the crucial difference between the three kinds of agents, defined according to their commitment strategies as discussed in Section 2, manifests itself in their reaction to different kinds of information received through different links. For all types of agents final revision of commitments takes place in the component intention_and_commitment_determination, namely in the components committed_goal_determination and committed_plan_determination. These are the components in which the knowledge about different commitment strategies resides.

To be more specific, the *blindly committed agent* only drops a committed_goal as a reaction to the receipt of information that the relevant goal has been realized. This information is transferred from the component belief_determination through the link transfer_belief_info_for_id, which in turn receives it through the link import_ws_info_for_bd, from the higher level components world_management and possibly from the component agent_specific_tasks. Some of the relevant generic knowledge present in the component committed_goal_determination is the following:

if own_commitment_strategy(blind) **and** goal_reached(X:Statement)
then to_be_dropped_committed_goal(X:Statement)

If this rule succeeds, an information link from committed_goal_determination to itself transfers the conclusion to_be_dropped_committed_goal(X:Statement) to update the atom inadequate_committed_goal(X:Statement) to true, which, in turn leads to the retraction of the committed goal, as described in Section 5.4.3. For simplicity these update links have not been depicted in Figure 4.

The *single-minded agent*, in addition, drops a committed_goal as a reaction to the information that the relevant goal can no longer be realized. This information is

transferred from the component belief_determination. The knowledge present in the component committed_goal_determination includes the following:

if own_commitment_strategy(single_minded) **and** goal_reached(X:Statement)
then to_be_dropped_committed_goal(X:Statement)

if own_commitment_strategy(single_minded) **and** goal_not_achievable(X:Statement)
then to_be_dropped_committed_goal(X:Statement)

The information goal_not_achievable(X:Statement), in turn, may depend on beliefs. In the first case the information may be transferred through the link import_ws_info_for_bd, from the higher level component world_management. In the second case plan revision is involved. In either case the relevant committed_plan is dropped using knowledge in the component committed_plan_determination:

if own_commitment_strategy(single_minded) **and** plan_not_achievable(X: plan)
then to_be_dropped_committed_plan(X: plan)

Next, in the second case, in order to check whether the relevant goal is achievable, the component plan_determination tries to design a plan. If this component succeeds in designing a new plan, this plan is adopted, and the original goal is maintained. If not, the component comes to the conclusion (based on exhaustive search) that no new plan can be designed. The component committed_goal_determination derives that the original goal must be retracted. Information specifying the success or failure of the design of a new plan is transferred from the component plan_determination to the component committed_goal_determination.

The *open-minded agent*, finally, in addition to the reasons adopted by the blindly committed agent and the single-minded agent, also drops a committed_goal in reaction to information that the goal is no longer desired, received from the component desire_determination through the link transfer_desire_info_for_id. The knowledge included in the component committed_goal_determination includes the following:

if own_commitment_strategy(open_minded) **and** goal_reached(X:Statement)
then to_be_dropped_committed_goal(X:Statement)

if own_commitment_strategy(open_minded) **and** goal_not_achievable(X:Statement)
then to_be_dropped_committed_goal(X:Statement)

if own_commitment_strategy(open_minded) **and** goal_not_desired(X:Statement)
then to_be_dropped_committed_goal(X:Statement)

In the last case the desire may have been dropped for many different reasons, not to be elaborated in this paper.

For all three agents, the stage of dropping a committed goal and/or a committed plan is followed by communication to the relevant agents. After this, a new committed goal should be established in the component intention_and_commitment_determination.

6.2 An example: meeting scheduling

To illustrate the use of explicit knowledge of dependencies and strategies for belief, intention and commitment revision, within the BDI model (specified within the DESIRE framework), a small, simplified example of an application, namely meeting scheduling, is described.

Three agents A1, A2 and A3 all believe that a meeting is required, and that their presence at this meeting is desired. They also believe that all three agents' presence is required. As agreement has been reached on a specific time slot, they all have an additional desire, namely to be at a meeting at the specific time slot.

The goal to be at a meeting in general, and at the specific meeting in particular, has been adopted by all three agents as an intended and committed goal. To accomplish this goal they all intend, and have committed to a plan to be at the specific meeting. In this example all three agents are single-minded. Below, the revision of attitudes is described from the point of view of A3. Agent A1 discovers that agent A2 is no longer available at the given time slot for the meeting.

Communication is required:
Agent A1 informs agent A3 of this fact.
As agent A3 believes that information A1 conveys is true, agent A3 also believes that agent A2 is no longer available.

Belief revision:
Given this new belief, agent A3 realizes that a prerequisite for the meeting (namely that all three participants' presence is required) no longer holds, and that the meeting can not be held as planned.

Dropping of committed goal:
As A3 is a single-minded agent, it is now allowed to drop its committed goal and the associated committed plan of meeting at the specific meeting.

Desire revision:
The desire to hold a meeting remains. The desire to hold the specific meeting is retracted.

Intention and commitment revision:
Agent A3's intention and commitment to the general goal of holding a meeting with the three other agents, still holds. Its intention and commitment to the goal of holding the specific meeting are retracted.
The intention and commitment to the plan for the specific meeting are also retracted.

The stage *Dropping of committed goal* follows the specification for single-minded agents elaborated in Section 6.1; the other stages can be described similarly (see (Brazier, Dunin-Keplicz, Treur and Verbrugge, 1997) for an extended specification).
In the example above, both committed and intended goals are dropped during intention and commitment revision. However, there are examples in which a committed goal is retracted while the corresponding intended goal remains; for example, a single-minded agent may become ill and retract its commitment to be present at the meeting, while still keeping its intention to be there (hoping to have recovered before the meeting).

7 Discussion and conclusions

In this paper a generic model for a rational BDI agent with explicit knowledge of dependencies between motivational attitudes has been modelled in DESIRE. The BDI model also includes knowledge of different commitment strategies in which these dependencies are used to reason about beliefs, desires and intentions, but also to explicitly revise specific beliefs, desires and/or intentions. Communication, action and observation may influence an agent's beliefs, desires, goals and plans dynamically.

The formal specification in DESIRE provides a bridge between logical theory, e.g. (Rao and Georgeff, 1991) and practice of BDI agents. Another bridge is described in (Rao, 1996), in which the operational semantics of a language corresponding to the implemented system dMARS, are formalized. Our model, in contrast, emphasizes the analysis and design methods of BDI systems, as do the architectures of (Jennings, 1995; Kinny, Georgeff and Rao, 1996). However, there are differences as well: our specification is more formal than Jennings' specification in (Jennings, 1995). DESIRE has a logical basis for which a temporal semantics has been defined (Brazier, Treur, Wijngaards and Willems, 1995). In contrast to the BDI architecture described in (Kinny, Georgeff and Rao, 1996), in our approach dynamic reasoning about beliefs, desires and goals, during plan execution, may lead to the construction of a (partially) new plan. This is partly caused by the parallel nature of specific reasoning processes in this model, but is also a consequence of the nature of explicit strategic knowledge of commitment strategies in the model. Strategic knowledge is used to revise, for example, beliefs, but also to revise intentions and commitments to goals and plans, during a dynamic process. Revisions are propagated by transfer of updated information on beliefs, desires and intentions to the components that need the information: components that reason about beliefs, desires, intentions, goals and plans.

The nature of continual activation of components and links makes it possible to transfer updated or new beliefs "automatically" to the relevant components. (The compositional revision approach incorporated in DESIRE is discussed in more depth in (Pannekeet, Philipsen and Treur, 1992)). In the paper the example of new information received from another agent, which may influence beliefs on which a goal has been chosen, is used to illustrate the effect this may have on the execution of a plan. Retraction of beliefs may lead to retraction of a number of goals that were based on these beliefs, which in turn may lead to retraction of a commitment to these goals. If the belief is the basis for a commitment to a plan, retraction of the belief may result in the retraction of the commitment to the plan and thus to its execution.

The DESIRE framework provides support in distinguishing the types of knowledge required to model rational agents based on mental attitudes. An existing agent architecture provided the basis for the model and the specification language provided a means to express the knowledge involved. By declaratively specifying task control knowledge and information exchange for each task, the dynamic process of revision has been explicitly specified.

The model as such provides a basis for further research: within this model more specific patterns of reasoning and interaction can be modelled and specified. Maintenance goals can be considered, joint commitments and joint actions can be modelled, more extensive communication patterns between agents can be analysed and represented, relative importance of intentions can be expressed, et cetera.

In contrast to general purpose formal specification languages such as Z and VDM, DESIRE is committed to well-structured compositional models. Such models can be specified in DESIRE at a higher level of conceptualisation than in Z or VDM and can be implemented automatically through use of automated implementation generators.

Acknowledgments

This work was partially supported by the Polish KBN Grants 3 P406 019 06 and 8T11C 03110.

References

Bratman, M.A. (1987). Intentions, Plans, and Practical Reason, Harvard University Press, Cambridge, MA.

Brazier, F.M.T. , Dunin-Keplicz, B., Jennings, N.R. and Treur, J. (1995). Formal specification of Multi-Agent Systems: a real-world case. In: V. Lesser (Ed.), Proc. of the First International Conference on Multi-Agent Systems, ICMAS-95, MIT Press, Cambridge, MA, pp. 25-32.

Brazier, F.M.T. , Dunin-Keplicz, B., Jennings, N.R. and Treur, J. (1997). DESIRE: modelling multi-agent systems in a compositional formal framework, International Journal of Cooperative Information Systems, M. Huhns, M. Singh, (Eds.), special issue on Formal Methods in Cooperative Information Systems: Multi-Agent Systems, vol. 6, to appear.

Brazier, F.M.T., Dunin-Keplicz, B., Treur, J., Verbrugge, R. (1997) A generic BDI architecture. Technical Report, Department of Mathematics and Computer Science Vrije Universiteit Amsterdam.

Brazier, F.M.T., Treur, J. (1996). Compositional modelling of reflective agents. In: B.R. Gaines, M.A. Musen (Eds.), Proc. of the 10th Banff Knowledge Acquisition for Knowledge-based Systems workshop, KAW'96, Calgary: SRDG Publications, Department of Computer Science, University of Calgary, pp. 23/1-13/12.

Brazier, F.M.T., Jonker, C.M., Treur, J., (1997). Formalisation of a cooperation model based on joint intentions. In: Proc. of the ECAI'96 Workshop on Agent Theories, Architures and Languages, ATAL'96. In: J.P. Muller, M.J. Wooldridge, N.R. Jennings, Intelligent Agents III, Lecture Notes in AI, vol. 1193, Springer Verlag, 1997, pp. 141-156.

Brazier, F.M.T., Treur, J., Wijngaards, N.J.E. and Willems, M. (1996). Temporal semantics of complex reasoning tasks. In: B.R. Gaines, M.A. Musen (Eds.), Proc. of the 10th Banff Knowledge Acquisition for Knowledge-based Systems workshop, KAW'95, Calgary: SRDG Publications, Department of Computer Science, University of Calgary, pp. 15/1-15/17

Castelfranchi, C. (1995). Commitments: From individual intentions to groups and organizations. In: V. Lesser (Ed.), Proc. of the First International Conference on Multi-Agent Systems, ICMAS-95, MIT Press, Cambridge, MA, pp. 41-48.

Cohen, P.R. and Levesque, H.J. (1990). Intention is choice with commitment, Artificial Intelligence 42, pp. 213-261.

Dennett, D. (1987). The Intentional Stance, MIT Press, Cambridge, MA.

Dunin-Keplicz, B. and Treur, J. (1995). Compositional formal specification of multi-agent systems. In: M. Wooldridge and N.R. Jennings, Intelligent Agents, Lecture Notes in Artificial Intelligence, Vol. 890, Springer Verlag, Berlin, pp. 102-117.

Dunin-Keplicz, B. and Verbrugge, R. (1996). Collective commitments. To appear in: Proceedings of the Second International Conference on Multiagent Systems, ICMAS-96.

Fagin, R., Halpern, J., Moses, Y. and Vardi, M. (1995). Reasoning about Knowledge. Cambridge, MA, MIT Press.

Jennings, N.R. (1995). Controlling cooperative problem solving in industrial multi-agent systems using joint intentions, Artificial Intelligence 74 (2).

Kinny, D., Georgeff, M.P., Rao, A.S. (1996). A Methodology and Technique for Systems of BDI Agents. In: W. van der Velde, J.W. Perram (Eds.), Agents Breaking Away, Proc. 7th European Workshop on Modelling Autonomous Agents in a Multi-Agent World, MAAMAW'96, Lecture Notes in AI, vol. 1038, Springer Verlag, pp. 56-71

Pannekeet, J.H.M., Philipsen, A.W. and Treur, J. (1992). Designing compositional assumption revision, Report IR-279, Department of Mathematics and Computer Science, Vrije Universiteit Amsterdam, 1991. Shorter version in: H. de Swaan Arons et al., Proc. Dutch AI-Conference, NAIC-92, 1992, pp. 285-296.

Rao, A.S. (1996). AgentSpeak(L): BDI Agents Speak Out in a Logical Computable Language. In: W. van der Velde, J.W. Perram (eds.), Agents Breaking Away, Proc. 7th European Workshop on Modelling Autonomous Agents in a Multi-Agent World, MAAMAW'96, Lecture Notes in AI, vol. 1038, Springer Verlag, pp. 42-55.

Rao, A.S. and Georgeff, M.P. (1991). Modeling rational agents within a BDI architecture. In: R. Fikes and E. Sandewall (eds.), Proceedings of the Second Conference on Knowledge Representation and Reasoning, Morgan Kaufman, pp. 473-484.

Shoham, Y. (1993). Agent-oriented programming, Artificial Intelligence 60 (1993) 51- 92.

Shoham, Y. (1991). Implementing the intentional stance. In: R. Cummins and J. Pollock (eds.), Philosophy and AI, MIT Press, Cambridge, MA, 1991, pp. 261-277.

Shoham, Y. and Cousins, S.B. (1994). Logics of mental attitudes in AI: a very preliminary survey. In: G. Lakemeyer and B. Nebel (eds.) Foundations of Knowledge Representation and Reasoning, Springer Verlag, pp. 296-309.

Tan, Y.H. and Treur, J. (1992). Constructive default logic and the control of defeasible reasoning, Report IR-280, Vrije Universiteit Amsterdam, Department of Mathematics and Computer Science, 1991. Shorter version in: B. Neumann (ed.), Proc. 10th European Conference on Artificial Intelligence, ECAI'92, Wiley and Sons, 1992, pp. 299-303.

Velde, W. van der and J.W. Perram J.W. (Eds.) (1996). Agents Breaking Away, Proc. 7th European Workshop on Modelling Autonomous Agents in a Multi-Agent World, MAAMAW'96, Lecture Notes in AI, vol. 1038, Springer Verlag.

Wooldridge, M. and Jennings, N.R. (1995). Agent theories, architectures, and languages: a survey. In: M. Wooldridge and N.R. Jennings, Intelligent Agents, Lecture Notes in Artificial Intelligence, Vol. 890, Springer Verlag, Berlin, pp. 1-39.